Supply Chain Strategy for Managing Risk for Health Insurance: An Application of Bayesian Model

Sukono[1], Endang Rusyaman[1], Jumadil Saputra[2*], Diana Ekanurnia[1], Yuyun Hidayat[3]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jawa-Barat, Indonesia*
[2]*School of Social and Economic Development, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia*
[3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor Jawa-Barat, Indonesia*
Corresponding author: jumadilsaputra@umt.edu.my

Abstract- **One important matter that should be investigated is the estimation of the risk distribution model of claims for each age group of the insured in health insurance because it is beneficial to prevent the occurrence of losses for insurance companies in the future. Supply chain strategy can be used in health insurance industry to manage the risks. In this paper, the research done is about the risk distribution model estimation on health insurance claims using Bayesian. The objective is to derive a health insurance risk model and determine the amount of net premium for each insured age group in health insurance. The sample of this study is the participant of health insurance in the Bandung area, Indonesia, especially for the insured who live in flood-prone areas. The estimation of the Poisson and Gamma distribution parameter is performed using the Bayesian method, which Open BUGS involves Markov Chain Monte Carlo (MCMC) the simulation technique. The estimation results show that the frequency of claims significantly follows the Poisson distribution, whereas the amount of claims substantially follows the Gamma distribution. With the result of the analysis, the estimated frequency distribution of claims and the amount of claims, a health insurance risk model may be established. Thus the net premium of health insurance for every age group, for the insured who live in the area prone to floods can be determined.**

Keywords- *Health insurance, Collective risk model, supply chain strategy, Bayesian modeling, Open BUGS.*

1. Introduction

Health is a valuable asset of human life. Everyone will certainly do everything possible to avoid any risks that may jeopardize their health, such as the possibility of suffering from a disease that requires huge costs. One way to avoid the risk of health financially, namely by being a member of health insurance [1]. Health insurance is a type of insurance products that guarantee medical care or the care of the insured when they have health problems or accidents. In the world of insurance, the parties who accept the risks are called to be the insurer or in this case are insurance companies, on the other side, the other parties who move the risk called to be insured risk [2]. When a person is insuring their health, then the person has transferred the financial risk, as a result of her health problems to the health insurance company. Because there is a transfer of financial risks, the insured has an obligation to pay a premium to the insurance company [3]. Amount of premium to be paid by policyholders to insurance companies are based on several factors, namely the insured age, medical history and selected protection plan [4]. In practice, it often that quite unnecessary cost are found and may cause financial losses for insurance companies. The effort that can be done by insurance companies to reduce unnecessary costs is by constructing a risk model which involves *uncertainty* that occurred during the loss process [5]. The risk model should be estimated by using certain method, to get a model that can be used to decide the premiums amount accurately. In [4], the health insurance model using Bayesian methods was analyzed. Similarly, [4], analyzed health insurance data also using the Bayesian method. Furthermore, [6] it creates a model of hierarchical health insurance claims using the Bayesian method. Another study [7], concern to the estimation model of the risk of life insurance claims for cancer patients using the Bayesian approach. For estimating the risk of the claim, the insurance participant data is grouped into two: the number of policies issued and the number of claims incurred, besides that, the model was used to

IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

estimate the risk value of life insurance claims each age group for each sex. Research [4], [8], [6], and [7], are estimating the health insurance risk model using Bayesian methods, significantly produce a model which can be used to determine the amount of health insurance claims and premiums. To conduct the estimation of the health insurance model using Bayesian methods, mostly done by using (MCMC) [8; 9]. [10], developed a model for aggregate losses on the evaluation of health insurance system premiums. In their research, they proposed the Bayesian approach in generalizing the collective risk model, and the prediction of aggregate losses was based on the Gibbs sampling algorithm. The objective is to obtain a predictive distribution of the aggregate losses in each class of the insured person. Based on the description above and apply the model proposed [10], this paper researched the assessment of the claim risk model for health insurance using Bayesian. The purpose of this study was to estimate the health insurance claims for people who live in the south Bandung area, West Java, Indonesia. The region is a concern in this study because it is often affected by flooding from the Citarum River annually, so their health are disturbed and proposed the health insurance claim often. This research is important to get a Mathematical model which can predict the amount of health insurance claims so that insurance companies do not suffer any losses. Also, it can be used to determine the amount of net premium of health insurance reasonably. Analysis of valuation models in this study was conducted using the Bayesian method because this method considered suitable to estimate the health insurance risk model.

2. Materials and Methods

2.1 Materials

The data used in this study is the health insurance data derived from an insurance company, which serves the participants of south Bandung area [20], West Java, Indonesia, from 2014 to 2016. Many data collection and the amount of health insurance claim, concerning research Yu [6]. Data collected consist of the insured population, $M_{a,t}$, the frequency of claims $N_{a,t}$, and the number of aggregate claim $X_{a,t}$, which categorized into seven groups of age and observed for twenty months. The grouping of data is based on the age of the insured, i.e. for those with aged 0 to 18 years (age group 1), ages 19 years to 35 years (age group 2), ages 36 to 45 years (age group 3), aged 46 to 55 years (age group 4), aged 56 to 65 years (age group 5), ages 66 to 75 years (age group 6), age over 75 years (age group 7).

2.2 Material and Methods

2.2.1 Bayesian Inference

In this section, a Bayesian inference study is conducted to estimate the model parameters of the claims frequency and the claims amount. Several steps need to be done to obtain the posterior of claims frequency, the aggregate of the claims, and the insured population. Posterior can be useful in renewing the parameters confidence level of the data.

2.2.2 Claim Frequency Model

The assumed frequency of claims $N_{a,t}$ is a random sample of discrete Poisson distribution with mean $\lambda_a m_{a,t}$ then:

$$N_{a,t} \sim \text{Poisson}\, \lambda_a M_{a,t}), \qquad a = 1,..,A; \quad t = 1,\ldots,T; \tag{1}$$

Where $M_{a,t}$ is an insured population in the age group of a over a period of time t and λ_a, is a parameter which specifies the number of examinations by the doctor to the insured at the age group of a [4; 10; 11]. Likelihood function of the data sample is given by:

$$l(n_{a,t} \mid \lambda_a m_{a,t}) \propto \lambda_a^{\sum_{t=1}^{T} n_{a,t}} e^{-\lambda_a \sum_{t=1}^{T} m_{a,t}},$$

Where $n_{a,t}$ the value of a random variable of $N_{a,t}$. Likelihood estimators of the parameter λ_a obtained by differentiating the logarithm of the likelihood function respect to λ_a and made to be equal to zero, in order to obtain:

$$\hat{\lambda}_a = \frac{\sum_{t=1}^{T} n_{a,t}}{\sum_{t=1}^{T} m_{a,t}}. \tag{2}$$

According to Ntzoufras [12] and Hamadu & Adeleke [13], the claim of frequency $N_{a,t}$ is Poisson distributed, and has a conjugate prior with Gamma distributed, expressed as:

$\lambda_a \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda)$, (3)

in which the probability density function (pdf) of Poisson distribution to claim frequency model is expressed as follows:

$$f(\lambda_a) \propto \lambda_a^{\alpha_\lambda - 1} e^{-\lambda_a \beta_\lambda}\ ;\ 0 \le \lambda_a < \infty .$$

Posterior to the parameter λ_a can be expressed by multiplying the likelihood $f(n_{a,t} \mid \lambda_a m_{a,t})$ by prior $p(\lambda_a)$ so that:

$$f(\lambda_{a,t} \mid n_{a,t} m_{a,t}) \propto \lambda_a^{\sum_{t=1}^{T} n_{a,t} + \alpha_\lambda - 1} e^{-\lambda_a (\sum_{t=1}^{T} m_{a,t} + \beta_\lambda)}$$

Because the posterior of the parameter λ_a is Gamma, so the mean of Gamma distribution can be obtained, given by:

$$\hat{\mu}_{\lambda_a} = E(\lambda_a \mid n_{a,t}) = \frac{\sum_{t=1}^{T} n_{a,t} + \alpha_\lambda}{\sum_{t-1}^{T} m_{a,t} + \beta_\lambda},$$

(4)

whereas the variance of posterior is given by:

$$\hat{\sigma}^2_{\lambda_a} = Var(\lambda_a \mid n_{a,t}) = \frac{\sum_{t=1}^{T} n_{a,t} + \alpha_\lambda}{(\sum_{t-1}^{T} m_{a,t} + \beta_\lambda)^2}.$$

(5)

If the prior uses, is a low-information prior then for Gamma distribution, the parameter which can be used are $\alpha_\lambda = \beta_\lambda = 10^{-3}$, where the lower the value of α_λ the greater the variance. Therefore, if $\alpha_\lambda = \beta_\lambda \to 0$ the mean posterior λ_a will be close to or equal to the mean sample $\hat{\lambda}_a$ [14-18]

2.2.3 Amount of Claim Model

The aggregate or total of the claim $X_{a,t}$ is assumed to be Gamma distributed with parameters $n_{a,t}$ and θ_a, are given by:

$X_{a,t} \sim \text{Gamma}(n_{n,t}, \theta_a)$,

(6)

where θ_a is the cost level of individual claims which is proposed by insured on the class age of a and $n_{a,t}$ is the number of claims proposed by insured in the age group of a over a period of time t [14]. Further, the aggregate of the claims that are estimated by likelihood function is given by:

$$l(x_{a,t} \mid n_{a,t} \theta_a) \propto \theta_a^{\sum_{t=1}^{T} n_{a,t}} e^{-\theta_a \sum_{t=1}^{T} x_{a,t}}.$$

(7)

Likelihood estimates of the parameter θ_a are obtained by differentiating the natural logarithm of likelihood function of equation (7) and made to be equal to zero, in order to obtain:

$$\hat{\theta}_a = \frac{\sum_{t=1}^{T} n_{a,t}}{\sum_{t=1}^{T} x_{a,t}}.$$

(8)

According to Migon and Penna [4], a parameter θ_a which is Gamma distributed has a conjugate prior with Gamma distributed with hyperparameter α_θ and β_θ can be expressed by:

$$\theta_a \sim \text{Gamma}(\alpha_\theta, \beta_\theta).$$

(9)

The probability density function of the parameter θ_a is given by:

$$g(\theta_a) \propto \frac{\beta_\theta^{\alpha_\theta} \theta_a^{\alpha_\theta - 1} e^{-\theta_a \beta_\theta}}{\Gamma(\alpha_\theta)}\ ;\ 0 \le \theta_a < \infty ,$$

Where $E(\theta_a) = \dfrac{\alpha_\theta}{\beta_\theta}$ and $Var(\theta_a) = \dfrac{\alpha_\theta}{\beta_\theta^2}$.

Posterior of the parameter θ_a is obtained by multiplying the likelihood $l(x_{a,t} \mid n_{a,t}, \theta_a)$ and prior $g(\theta_a)$ can be expressed as follows:

$$g(\theta_a \mid n_{a,t} x_{a,t}) \propto \theta_a^{\sum_{t=1}^{T} n_{a,t} + \alpha_\theta - 1} e^{-\theta_a (\sum_{t=1}^{T} x_{a,t} + \beta_\theta)}$$

Because the posterior of the parameter θ_a is Gamma distributed, then the mean of Gamma distribution can be obtained which expressed by:

$$\hat{\mu}_{\theta_a} = E(\theta_a \mid x_{a,t}) = \frac{\sum_{t=1}^{T} n_{a,t} + \alpha_\theta}{\sum_{t-1}^{T} x_{a,t} + \beta_\theta},$$
(10)

whereas the variance of posterior given by:

$$\hat{\sigma}^2_{\theta_a} = Var(\theta_a \mid n_{a,t}) = \frac{\sum_{t=1}^{T} n_{a,t} + \alpha_\theta}{(\sum_{t-1}^{T} x_{a,t} + \beta_\theta)^2}.$$
(11)

The mean should be calculated to determine the location of the posterior distribution of the number line. Meanwhile, the variance is determined to portray the posterior distribution of that distribution [8; 15].

Assumed that the amount of individual claim is following an exponential distribution with the mean $1/\theta_a$ for $a = 1,..,A$, then:

$$Z_{a,t} \mid \theta_a \sim \text{Expenential}(1/\theta_a); \qquad \theta_a > 0.$$
(12)

Where θ_a is the level of personal claim cost, which by agreement is paid by insurance companies to the insurer for each examination with the doctor [4]? The probability density function of the parameter $Z_{a,t,i}$, can be expressed by:

$$g(Z_{a,t} \mid \theta_a) = \frac{1}{\theta_a} e^{-\frac{z_{a,t}}{\theta_a}}, \qquad (13)$$

where $E(Z_{a,t} \mid \theta_a) = \frac{1}{\theta_a}$ and $Var(Z_{a,t} \mid \theta_a) = \frac{1}{\theta_a^2}$.

2.2.4 Health Insurance Risk Model

Given a policy portfolio classified by age group of $a = 1,..,A$ which is observed over a period of time $t = 1,...,T$. In Migon and Penna [4], it is supposed that $N_{a,t}$ is representing a claim frequency, $X_{a,t}$ is representing the aggregate of the claim, and $Z_{a,t,i}$ is representing the amount of i-th individual claim which proposed in the interval of time $(t-1,t)$ in the age group of a. Collective risk model can be expressed as follows:

$$X_{a,t} = \sum_{i=1}^{N_{a,t}} Z_{a,t,i}; \; a = 1,..,A; \; t = 1,...,T.$$
(14)

The assumptions used in this model are:

- The amount of individual claim $Z_{a,t,i}$ is a non-negative random variable which is independent and identically distributed.
- Claim frequency $N_{a,t}$ is a random variable that independent one another respect to claim amount of i-th, that is $Z_{a,t,i}$.

The advantage of the collective risk model is a frequency of claim and amount of claim can be modeled separately. The population of insured $M_{a,t}$ is time-varying for a known age group of a, then the collection of the number of aggregate claims $\{X_{a,t}: \; a = 1,..,A, \; t = 1,...,T\}$non-identically distributed [17].

If $n_{a,t} = \lambda_a m_{a,t}$ with the value of a and t are constant, then the claim amount of $Z_{a,t,i}$ for $i = 1,2,...$ age group of a at the period of time t is independent and identically distributed. Further, the expectation and variance of the amount of aggregate claim can be obtained by multiplying the claim frequency of expectation with i-th the amount of claim expectation, so that:

$$\frac{E(X_{a,t})}{m_{a,t}} = \frac{1}{m_{a,t}}\left(E\{E(X_{a,t} \mid N_{a,t})\}\right)$$

$$= \frac{1}{m_{a,t}}\left(E(N_{a,t})E(Z_{a,t,i})\right) = \frac{\lambda_a}{\theta_a},$$
(15)

and

$$\frac{Var(X_{a,t})}{m_{a,t}} = \frac{1}{m_{a,t}}\left(E\{Var(X_{a,t} \mid N_{a,t})\} + Var(E\{X_{a,t} \mid N_{a,t}\}\right)$$

$$= \frac{1}{m_{a,t}}\left(E(N_{a,t})Var(Z_{a,t,i}) + \{E(Z_{a,t,i})\}^2 Var(N_{a,t})\right)$$

$$= \frac{1}{m_{a,t}} \left(\frac{2\lambda_a m_{a,t}}{\theta_a^2} \right) \propto \frac{\lambda_a}{\theta_a^2},$$
(16)

Where $n_{a,t} = \lambda_a m_{a,t}$ show the number of claims proposed by the insured population at the age group of a at the period of time t. That expectation and variance can be used to determine a pure premium which should be paid by the insured to the insurance company [19].

2.2.5 Open BUGS

Open BUGS is a software application for the Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods. Open BUGS is the open source variant of Win BUGS (Bayesian inference Using Gibbs Sampling). It runs under Windows and Linux, as well as from inside the R statistical package. Versions from v3.0.7 onwards have been designed to be at least as efficient and reliable as Win BUGS over a range of test applications [12; 14; 21 ;22].

3. Results

This section analyzes the results, which include: the model parameters of the frequency of claims, the claim amount, claim aggregate, and the determination of net premiums. A parameter value of claim frequency and a number of claim on the health insurance risk models need to be determined to calculate the cost average of net premium that should be paid by the insured to a particular age group to the insurance company.

3.1 Data Analysis of Claim Frequency

In health insurance, claims frequency can be significantly affected by the number of examination to the doctor performed by the insured in the age group of a each month λ_a. In equation (1), the claim frequency filed by the age group of a in the period of time t $N_{a,t}$ is assumed to be Poisson distributed, with λ_a as the parameter. Based on the data being analyzed, the estimation value of the likelihood of parameter λ_a determined by using equation (2), as follows:

$$\hat{\lambda}_a = \frac{1270}{6860} = 0.18513 \qquad (17)$$

Where $\hat{\lambda}_a$ is the average level of a number of claims from the insured in the age group of a for each month? In this study, it is assumed that the prior distribution for the parameter λ_a is Gamma with hyperparameter α_λ =0.01 and β_λ = 0.01 expressed by:

$$\lambda_a \sim \text{Gamma}(0.01, 0.01).$$
(18)

The value of α_λ and β_λ are used in the computation process by OpenBUGS program. Based on equation (4), the posterior distribution of parameter λ_a is Gamma distributed with hyper parameter $\sum_{t=1}^{T} n_{a,t} + \alpha_\lambda$ and $\sum_{t=1}^{T} m_{a,t} + \beta_\lambda$.

Because the data frequency of claims and the insured population is known, as well as the value of α_λ and β_λ have been determined, then the posterior distribution of parameter λ_a can be expressed by:

$$\lambda_a \mid n_{a,t} \sim \text{Gamma}(1270.01, 6860.01),$$
(19)

Where the expectation $\hat{\mu}_{\lambda_a}$ can be determined by equation (4), the result is $\hat{\mu}_{\lambda_a}$ = 0.18513. While the variance $\hat{\sigma}^2_{\lambda_a}$ can be determined based on the equation (5), the result is $\hat{\sigma}^2_{\lambda_a}$ = 2.7×10^{-5}, to obtain standard deviation $\hat{\sigma}_{\lambda_a}$ = 5.2×10^{-3}.

The average frequency of claims in each group of age can be obtained by determining its level of a number of examining to the doctor multiplied by the number of insured population per month for each group of age. OpenBUGS program used to obtain a statistical summary of the parameter λ_a by simulating data samples from the posterior distribution in several iterations. These simulations of the sample are based to Monte Carlo Markov Chain (MCMC) with Gibbs Sampler algorithm in which the initial values of the parameter λ_a specified in OpenBUGS program. In this study, three chains are used to simulate the sample of the posterior distribution, in which each chain was observed to test the convergence of a trace plot visually. The iteration that performed on each chain is of 10000 iterations with an initial value $\lambda_a^{(0)}$ equal to 1 that obtained from OpenBUGS program. Statistical summary by OpenBUGS program produces an estimate Bayes OpenBUGS program with parameters λ_a consisting of the mean, standard deviation, and 95% confidence intervals are presented in following Table 1.

Table1. The results of Bayesian estimation with parameters λ_a by *Open BUGS program*

λ_a	Mean	Standard Deviation	2.5% Percentile	97.5% Percentile
λ_1	0.2211	0.01499	0.1927	0.2515
λ_2	0.1693	0.01318	0.1445	0.1963
λ_3	0.1367	0.01185	0.1144	0.1608
λ_4	0.1744	0.01335	0.1494	0.2014
λ_5	0.1775	0.01350	0.1522	0.2052
λ_6	0.1705	0.01332	0.1457	0.1977
λ_7	0.2436	0.01566	0.2137	0.2751

Table 1 shows the mean numbers of claims from the insured in the group of age 1 approved by the insurance company every month, that is 0.2211, are obtained, it means that from 100 insured, the number of claims to be approved are 22 people. In the group of age 3, it is obtained that the mean number of approved claims by the insurance company is 0.1367 monthly, means that from 100 insured, the number of approved claims are 13 people. Two types of plots can be used to describe the simulation of the posterior distribution sample parameter λ_a that observed for twenty months, namely, trace plot and density plot. Trace plot is used to test the convergence of the posterior parameter simulation λ_a with MCMC method, in which the trace plot is said to be converged if it is used more than one chain, and those chains are overlapping each other in the same plot. Density Plot is used to describe the results of each simulation value of posterior samples in several iterations.

3.2 Data Analysis of Claim Amount

The total amount of claims is assumed as Gamma distributed with parameters $n_{a,t}$ and θ_a, with $n_{a,t}$ as a claim frequency, of which data are known, while the number of individual claims is assumed to be exponentially distributed with the parameter θ_a. Because the data of claim frequency and the total number of claims are known then the estimation value of likelihood with parameter θ_a can be expressed follows:

$$\hat{\theta}_a = \frac{1270}{4133354778} = 3.07256 \times 10^{-6},$$

(20)

Which means the average number of $i-th$, $i=1,\ldots,N_{a,t}$ claim filed by the insured for each group of age $a=1,\ldots,7$ per month is IDR 325,461.5044, stated by:

$$E(Z_{a,i,20}) = \frac{1}{3.07256 \times 10^{-6}} = 325{,}461.5044.$$

Prior distribution used for the parameter θ_a is a Gamma-distributed conjugate prior with hyperparameter α_θ and β_θ. Previously to parameters, θ_a can be assumed to have a hyperparameter value $\alpha_\theta = 0.01$ and $\beta_\theta = 0.01$ with $E(\theta_a) = 1$ dan $Var(\theta_a) = 100$. Based on the equation (10), expectations of the following parameters θ_a can be expressed by:

$$\hat{\mu}_{\theta_a} = \frac{1{,}270.01}{413{,}335{,}478.01} = 3.072259 \times 10^{-6}.$$

Statistical summary of the following parameter θ_a is shown in Table 2 which consists of the mean, standard deviation, and 95% credible interval for the parameters θ_a. The results are obtained by

OpenBUGS program of three parallel chains with an initial value $\theta_a^{(0)} = 1$ and the iterations done of 10,000 times.

The initial value $\theta_a^{(0)}$ is 1, given to parameter θ_a as an initial value to simulate the sample of posterior distribution by using MCMC algorithm. Three parallel chains with an iteration of 10,000 times on each chain are used to test the convergence of the parameter θ_a which examined by visually using trace plot. The convergence test is done by seeing whether the three chains are overlapping each other or not in the trace plot. The vertical axis of the trace plot represents the value of each sample, and the horizontal axis represents the number of iterations.

Table2. Statistical Summary of parameter Bayesian Estimation θ_a by OpenBUGS Program

θ_a	Mean ($\times 10^{-6}$)	Standard Deviation ($\times 10^{-7}$)	2,5% Percentile ($\times 10^{-6}$)	97,5% Percentile ($\times 10^{-6}$)	$E(Z_{i,a,20})$
θ_1	3.402	2.306	2.964	3.869	293,944.7
θ_2	2.688	2.093	2.294	3.117	372,023.8
θ_3	2.997	2.598	2.508	3.527	333,667.0
θ_4	3.093	2.367	2.649	3.572	323,310.7
θ_5	2.687	2.044	2.304	3.106	372,162.3
θ_6	3.255	2.542	2.78	3.774	307,219.7
θ_7	3.342	2.148	2.933	3.774	299,222.0

Based on Table 2, it is obtained that the number of the individual claim approved by the insurance company paid to the insured in the grouping of the age group of 2 is IDR 372,023.8 monthly. In the age group of 6, it is obtained that the mean number of the individual claim approved by the insurance company to be paid to the insured with the age group of 6 is IDR 307,219.66 monthly.

The primary purpose of this paper is to determine the net premium to be paid by the insured to the insurance company. Every premium depends on the value of claim frequency and the amount of claims. Net premiums can be determined by multiplying the result of the mean claim frequency and the amount of claims. Expectation and variance of the net premium by using the Bayesian method for each group of age a per month are stated in Table 3.

3.3 Estimation of Net Premium

Table3. The Prediction of the net premium per insured in the group of age $a = 1,\ldots,7$

Age Group (AG)	$\hat{\lambda}_a$	$\hat{\theta}_a$	Mean	Standard Deviation
AG 1	0.2211	0.0000034	64,991.18	138,216.60
AG 2	0.1693	0.0000027	62,983.63	153,073.20
AG 3	0.1367	0.0000030	45,612.28	123,366.60
AG 4	0.1744	0.0000031	56,385.39	135,018.50
AG 5	0.1775	0.0000027	66,058.80	156,794.70
AG 6	0.1705	0.0000032	52,380.95	126,856.10
AG 7	0.2436	0.0000033	72,890.48	147,683.60

Table 3 displays the estimation of net premium that must be paid by every insured to the insurance company in the age group of 3 amounted to IDR 45,612.28, while the estimation of net premium that must be paid by every insured in an insurance company in the age group of 6 has amounted to IDR 52,380.95. The insurance company can use the net premium estimation to determine the backup-cost that should be prepared so that the insurer can be avoided from making losses.

4. Discussion

Some insurance companies often refuse the participation of health insurance programs for the insured live in flood-prone areas. In fact, on the contrary, the government always advocates the role of insurance companies in lightening the burden of health insurance costs for people living in flood-prone areas. Rejection by some insurance companies is due to the collection of premiums obtained from the insured live in flood-prone areas cannot cover the amount of risk claims filed. Therefore, we try to evaluate the determination of the amount of health insurance premiums, especially for the insured live in flood-prone areas. In evaluating, we do by referring to some research results that have been done by other researchers. Among them are Amin and Salem10, which proposes the use of Bayesian methods in evaluating the determination of health insurance premiums. In evaluating the determination of health insurance premiums for insured persons living in these flood-prone areas, we use the Bayesian method, where the claim frequency distribution is assumed to follow the Poisson distribution having the conjugate prior distributed Gamma, and the amount of claims assumed to follow the Gamma distribution with conjugate prior also distributed Gamma. The Bayesian method was used in this study, as the Bayesian method has been widely used by researchers in estimating the parameters of a distribution. The Bayesian method is very well used especially for very complex distribution functions, or in other words, for the distribution function, it has more than two parameters. Also, the Bayesian method in parameter estimation incorporates the information contained in the sample, with other information previously available. In terms of classical assumptions, it is assumed that the population parameter has a certain unknown price so that the probability statement about the population parameter has no meaning. To evaluate the determination of health insurance premium, in this study we use health insurance data obtained from an insurance company in West Java, Indonesia. Data include the frequency of claims and the amount of claims during the period 2014-2016, and some information that cannot be obtained from insurance companies, we use information obtained from previous researchers. From this set of data, we divide into seven age groups, as described in the materials section. Furthermore, each age group is estimated the amount of pure premium, as the basis for the determination of the premium by the insurance company. For estimating the amount of pure premium, it is necessary to estimate the parameters of the claim frequency distribution and the distribution parameter of the size of the claim. The frequency of claims can be affected by the number of visits to doctors for health care, by the insured in the age group a every month. The number of visits to the doctor by the insured is assumed to be Poisson distributed with parameters λ_a. It is assumed that the prior distribution is for the parameter λ_a is Gamma with hyper-parametric α_λ =0.01 and β_λ = 0.01. Values α_λ and β_λ subsequently used in the estimation of posterior distribution parameters λ_a through the computing process using the Open BUGS program. Results of parameter estimation λ_a for seven age groups are given in Table 1. Based on Table 1, it can be shown that the insured who is rarely expected to visit a doctor is age group 3, i.e. the insured aged between 36 to 45 years, as much as 0.1367 times the number of the insured each month. Meanwhile, the most frequent insured to visit a doctor is a group of age 7, i.e. the insured who is more than 75 years old, as much as 0.2436 times the number of the insured every month. The amount of the claim can be affected by the cost of each visit to the doctor for health care, by the insured in the age group a every month. The magnitude of the assumption is assumed to be exponentially distributed with parameters $1/\theta_a$, where conjugate prior for parameters θ_a is distributed Gamma with hyper-parameter $\alpha_\theta = 0.01$ and $\beta_\theta = 0.01$. Values α_θ and β_θ subsequently used in the estimation of posterior distribution parameters θ_a through the computing process using the Open BUGS program. Results of parameter estimation θ_a for the seven age groups are given in Table 2. Based on Table 2, it can be shown that the age group which charges the lowest per one time claim is the insured between 0 and 18 years old or age group 1, amounting to IDR 293,944.70. Meanwhile, the biggest claim per one time claim submitted per month is the insured between the ages of 56 to 65 years or group 5, amounting to IDR 372,310.70.

Using the values in Tables 1 and 2 then used to determine the value of expansion and variance of aggregate claims in each age group. The cost of aggregate claims depends on the frequency of claims and the amount of claims of each age group. Based on the expected value and the variance of aggregate claims are used to determine the net premium for each age group. Net premiums can be determined by multiplying the result of the claim claims frequency and the amount of claims. The net premium calculation results are given in Table 3. Based on the aggregate claims value (mean) in Table 3, it is seen that the largest net premium to be paid by the insured person is in the age group 7, or the insured is more than 75 years old, amounting to IDR 72,890.48. It indicates that the insured, is more than 75 years old, has a risk of making claims at a higher cost than other age groups. The values of pure premium calculations in Table 3, of course, will be very useful for insurance companies, in determining health insurance premiums for the insured who live in flood-prone areas. Furthermore, the amount of the premium determined for each age group, depends also on the loading factor set by the insurance company. Loading factor such as how much percent of administrative costs, agency fees, and so forth but that also needs to be considered is the premium set by a competitor company, and also the level of people's purchasing power.

5. Conclusion

This study is an application of Bayesian approach to determine the amount of health insurance premium for the insured living in flood-prone areas, where the claim frequency is assumed to have a Poisson distribution which has conjugate prior distributed Gamma, while the amount of claim is assumed to have Gamma distribution with conjugate prior also Gamma distribution based on the supply chain strategy. This study will help insurance companies to determine the amount of health insurance premiums in each age group for the insured who live in flood-prone areas. This Bayesian approach can also be applied to the determination of the amount of other insurance premiums, caused by the occurrence of natural disasters. In this study, we have analyzed the estimation of health insurance risk model using Bayesian method. Based on the results of the analysis, it can be concluded that the frequency of claims following the Poisson distribution, while the magnitude of the claim follows the Gamma distribution. The health insurance risk model is based on Poisson and Gamma distribution, which is then used to determine the net premium of health insurance, especially for the insured living in flood-prone areas of Citarum Bandung Indonesia. The amount of net insurance premium for each age group of the insured is as follows: age 0 to 18 years of IDR 64,991,18s; ages 19 to 35 years of IDR 62,983.63; age 36 to 45 years of IDR 45,612.28; age 46 to 55 years of IDR 56,385.39; age 56 to 65 years of IDR 66,058.80; age 66 to 75 years of IDR 52,380.95; and age more than 75 years of 72,890.48. The estimated premium amount is beneficial for the insurance company as a consideration in determining the net premium rate for every age group of the insured. So the price that is set by the insured is affordable, and also does not cause loss to the insurance company.

Acknowledgments

Acknowledgments conveyed to the Rector, Director of Directorate of Research, Community Involvement and Innovation, and the Dean of Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, with whom the Internal Grant Program of Universitas Padjadjaran was made possible to fund this research. The grant is a means of enhancing research and publication activities for researchers at Universitas Padjadjaran.

References

[1] Aizcorbe, A., Liebman, E., Pack, S., Cutler, D.M., Chernew, M.E., and Rosen, A.B. *Measuring the health care costs of individuals with employer-sponsored health insurance in the U.S.: A comparison of survey and claims data*. Statistical Journal of the IAOS, 28, pp: 43–51, 2012.

[2] Bhuyan, S. and Baruah, R.D. *Health Care Expenditure Pattern of the Labour Class in the Organized and Unorganized Sector of Guwahati Role of Health Insurance as a Protective Mechanism*. EPRA International Journal of Research and Development (IJRD), 1(4), pp: 42-51, 2016.

[3] Sukmawan, Y.I.N., Husaini, and Marlinae, L. *Factors Related to the Non Specialist Referral Rates in Primary Health Care Kotawaringin Timur District*. EPRA International Journal of Research and Development (IJRD), 1(4), pp: 87-93, 2016.

[4] Migon, H.S. and Penna, M.O.P. *Bayesian Analysis of a Health Insurance Model*. Journal of Actuarial Practice, 13, pp: 61-80, 2006.

[5] Mesike, G.C., Adeleke, I.A., and Ibiwoye, A. *Predictive Actuarial Modeling of Health Insurance Claims Costs*. International Journal of Mathematics and Computation, 14(1), pp: 34-45, 2012.

[6] Yu, G.Q. *Hierarchical Bayesian Modelling of Health Insurance Claims. Master Thesis. Department of Statistics and Actuarial Science*, Faculty of Science, Simon Fraser University, 2015.

[7] Sukono, Suyudi, M., Islamiyati, F., and Supian, S. *Estimation model of life insurance claims risk for cancer patients by using a Bayesian method. IOP Conference Series*: Materials Science and Engineering, 166, 2017. DOI:10.1088/1757-899X/166/1/012022.

[8] Andrieu, C., Doucet, A., and Holenstein, R. *Particle Markov Chain Monte Carlo Methods*. Journal of the Royal Statistical Society. B, 72(3), pp: 269–342, 2010.

[9] Martin, A.D., Quinn, K.M., and Park, J.H. *MCMCpack: Markov Chain Monte Carlo in R*. Journal of Statistical Software, 42(9), 2011. DOI: 10.18637/jss.v042.i09.

[10] Amin, Z. and Salem, M. *Bayesian Modelling of Health Insurance Losses*. Journal of Applied Statistics, Volume 42, 2015, Issue 2, pp: 231-251, 2015. https://doi.org/10.1080/02664763.2014.947247

[11] Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., and Wagenmakers, E.J. *An Introduction to Bayesian Hypothesis Testing for Management Research*. Journal of Management, 41(2), pp: 521–543, 2015.

[12] Ntzoufras, I. *Bayesian Modeling Using WinBUGS. New Jerse*y: Wiley Series in Computational Statistics, 2009.

[13] Hamadu, D. and Adeleke, I. *Model-Assisted Credibility Rating for Health Insurance Claims*. Journal of Mathematics and Technology, 3(2), pp: 37-43, 2012. DOI: 10.7813/jmt.2012/3-2/7.

[14] Katsis, A. and Ntzoufras, I. *Bayesian Hypothesis Testing for the Distribution of Insurance Claim Counts Using the Gibbs Sampler*. Journal of Computational Methods in Sciences and Engineering, 5(3), pp: 201–214, 2005.

[15] Karnon, J., Mackay, M., and Mills, T.M. *Mathematical Modelling in Health Care. Working Paper*. 18th World IMACS/MODSIM Congress, Cairns, Australia 13-17 July 2009. Available from: http://mssanz. org.au/modsim09.

[16] Craiu, R.V. and Rosenthal, J.S. Bayesian *Computation via Markov chain Monte Carlo. Working Paper*. (Version of June 6). Department of Statistics, University of Toronto, 2013.

[17] Sukono, Nahar, J., Putri, F.T., Subiyanto, Mamat, M., and Supian, S. *Estimation of Outstanding Claims Reserving Based on Inflation Risk on Car Insurance Companies by Using the Bootstrap Method*. Far East Journal of Mathematical Sciences (FJMS). Volume 102, Number 4, pp. 687-706, 2017. ISSN: 0972-0871. http://dx.doi.org/10.17654/MS102040687.

[18] Xie, X., Zhang, G., Huang, Y., and Ou, S. *Bayesian Models for Healthcare Data Analysis*. Austin Journal of Biomedical Engineering, 1(3), pp: 1-4, 2014.

[19] Sukono, Hidayat, Y., Suhartono, Sutijo, B., Bon, A.T.B., and Supian, S. *Indonesian Financial Data Modeling and Forecasting by Using Econometrics Time Series and Neural Network*. Global Journal of Pure and Applied Mathematics. Volume 12, Number 4, pp. 3745–3757, 2016. ISSN 0973-1768. © Research India Publications. http://www.ripublication.com/gjpam.htm.

[20] Sidi, P., Mamat, M., Sukono, and Supian, S. *Evaluation Model for Risk Insurance Premium of Building Damage Caused by Flood: Case Study in Citarum Watershed, Southern Bandung, Indonesia*. Journal of Engineering and Applied Sciences, 12 (17): pp. 4420-4425, 2017. ISSN: 1816-949X. ©Medwell Journal, 2017.

[21] Sukono, Riaman, Lesmana, E., Wulandari, R., Napitupulu, H., and Supian, S. *Model Estimation of Claim Risk and Premium for Motor Vehicle Insurance by Using Bayesian Method. IOP Conf. Series*: Materials Science and Engineering, 300 (2018) 012027, pp. 1-10. doi:10.1088/1757-899X/300/1/012027.

[22] Sturtz, S., Ligges, U., and Gelman, A. *R2OpenBUGS: A Package for Running OpenBUGS from R*. Journal of Statistical Software, 12(3), pp. 1-16, 2005.

[23] Meysam, Davoodabadi Farahani, Hamid, Aghajani. *Identification of Potential Groundwater Zones Using RS and GIS*, UCT Journal of Research in Science , Engineering and Technology, Issue 4,pp.04-06, 2013.